

Distribution of k -Hop Paths in the Random Connection Model

Alexander P. Kartun-Giles, *Member, IEEE*, Sunwoo Kim, *Member, IEEE*

Abstract

We study the probability of k -hop connection between two nodes in a wireless multi-hop network, addressing the difficulty of providing an exact formula for the scaling of hop counts with Euclidean distance without first making a sort of mean field approximation, which in this case assumes all nodes in the network have uncorrelated degrees. We therefore study the mean and variance of the number of k -hop paths between two vertices x, y in the random connection model, which is a random geometric graph where nodes connect probabilistically rather than according to a law of intersecting spheres. In the example case where Rayleigh fading is modelled, the variance of the number of three hop paths is in fact composed of four separate decaying exponentials, one of which is the mean, which decays slowest as $\|x - y\| \rightarrow \infty$. These terms each correspond to one of exactly four distinct sub-structures which can form when pairs of paths intersect in a specific way, for example at exactly one node. We also discuss a potential application of our results for connectivity-based localisation.

Index Terms

Random connection model, random geometric graphs, wireless multi-hop networks, localisation, data capacity, route discovery.

I. INTRODUCTION

Take a homogeneous Poisson point process $\mathcal{Y} \subset [0, \sqrt{n}]^2$ with $\mathbb{E}\|\mathcal{Y}\| = n$, and form a graph by adding an edge between pairs of \mathcal{Y} whenever they are within Euclidean distance r_0 of each other, denoted $\mathcal{G}(n, \pi r_0^2(n))$. According to Penrose [1] and later Gupta and Kumar [2], if $\pi r_0^2(n)$

A. P. Kartun-Giles and S. Kim are with the Department of Electronics and Computer Engineering, Hanyang University, Seoul 11223, South Korea (e-mail: alexander@hanyang.ac.kr; remero@hanyang.ac.kr).

grows logarithmically with n , then G is connected (a path exists between every pair of nodes) with probability one as $n \rightarrow \infty$, since in that limit,

$$P(\mathcal{G}(n, \log n + c(n)) \text{ is connected}) \rightarrow e^{-e^{-c(n)}}. \quad (1)$$

and then one introduces any function $c(n) \rightarrow \infty$. Also in the limit $n \rightarrow \infty$, nodes which find themselves isolated do so independently with probability $\exp(-\pi r_0^2)$, since this becomes rare. In fact, the number of isolated vertices as a proportion of all vertices in the graph goes to zero, and a homogeneous Poisson point process of isolated vertices \mathcal{Y}_{n_0} is observed. Finally, with the added condition of logarithmic scaling as in Eq. 1, the probability that \mathcal{G} is disconnected but free of isolated nodes tends to zero i.e. any two ‘large’ clusters eventually merge. The point at which isolated nodes disappear is then precisely the same as the connectivity transition, with high probability (w.h.p., with probability one as $n \rightarrow \infty$). For the proof of this, see e.g. the section on connectivity in Walter’s review [3]. Similarly, in the random subgraph of the complete graph on n nodes obtained by including each of its edges independently with probability $p \sim \log n/n$, the probability that this graph is disconnected but free of isolated nodes tends to zero [4]. An interesting observation in e.g. the geometric case, but very much in both cases, is that even with the node degrees going to infinity with n , subcritical growth of πr_0^2 results in an infinite sea of isolated vertices amongst an otherwise connected graph, whereas extension to supercritical growth results in these vertices all suddenly merging into the giant component, which is a phase transition to connectivity.

As a sort of mix between the previous two examples, consider a *random* connection model of a set of nodes selected from a d -dimensional space, known as *soft random geometric graph* [5]–[8] when confined to a bounded region, or just *random connection model* [9]–[12] when the nodes are a countably infinite subset of \mathbb{R}^d . Take a Poisson point process $\mathcal{Y} \subset [0, 1]^d$ of intensity $\lambda(n)dx$, dx Lebesgue measure on \mathbb{R}^d , and $(\lambda(n))_{n \in \mathbb{N}}$ an increasing $(0, \infty)$ -valued sequence which goes to ∞ with n . Also, take the measurable function $H : \mathbb{R}^+ \rightarrow [0, 1]$ to be the probability that two nodes are joined by an edge. Then, as $\lambda(n) \rightarrow \infty$ along this sequence, in any limit where the expected number of isolated nodes converges to a positive constant $\alpha < \infty$, i.e.

$$\lambda \int_{[0, \sqrt{n}]^2} \exp \left(-\lambda \int_{[0, \sqrt{n}]^2} H(\|x - y\|) dy \right) dx \rightarrow \alpha, \quad (2)$$

their number converges to a Poisson distribution with mean α , see Theorem 3.1 in Penrose’s recent paper [5]. The connection probability then follows, as before, from the probability that

the graph is free of isolated vertices, given some conditions on the rate of growth of H with n , see e.g. [5], [6] for the case of random connection in a confined geometry, or non-convex geometry [7], or for the random connection model [11].

These scaling laws for both deterministic and probabilistic connection are central to the theory of random geometric graphs. Other network observables also scale in space, with data capacity a well cited example [13]. Practically important and related to both is the length $k \in \mathbb{N}$ of the multi-hop paths which run through the network in e.g. the connectivity regime of Eq. 1. Bounds on the distribution of the number of hops between two points in space, for example, has been a recent focus of many researchers interested in the statistics of the number of hops to e.g. a sink in a wireless sensor network, or gateway-enabled small cell in an ultra-dense deployment of non-enabled smaller cells [11], [14]–[16], since it relates to data capacity in e.g. multihop communication with infrastructure support [17], [18], route discovery [19], [20] and localisation [21], among other problems where the thermodynamic or dense limits of random geometric graphs are relevant or applicable. For recent work on this, see e.g. Díaz, Mitsche, Perarnau and Pérez-Giménez [22] for the case of deterministic connection, or [10] for the case of the random connection model, and the references therein. As noted, the limit where expected node degrees diverge does not necessarily imply anything about hop details, so studying the scaling required to e.g. have all nodes, perhaps at a bounded displacement, connect in k -hops or fewer, is a similar activity to previous work concerning the probability of asymptotic connection.

The consensus is that spatial dependence between the node degrees, which form a Markov random field [23] in the deterministic, finite range case, appears to preclude the exact description of a map between distances, given by some norm, and the space of distributions describing the probability of k -hop connection between two nodes of known displacement, see e.g. Section 1 of [15]. We believe, however, that the number $\sigma_k(\|x - y\|)$ of k -hop paths may have a probability generating function similar to a q -series common in other combinatorial enumeration problems [24]. In fact, the first author has demonstrated that this is indeed the case under deterministic connection in one dimension, proving that $\mathbb{E}q^{\sigma_k}$ is a random q -multinomial coefficient [25]. In this paper, and in order to eventually provide corrections to what is a sort of *mean field approximation* to this k -hop connectivity problem [26], we therefore derive $\text{Var}(\sigma_k(\|x - y\|))$, since this is the most tractable statistic which contains useful information about how paths in a random geometric graph influence each others existence. This has recently been studied in the deterministic case, but restricted, again, to one dimension as part of an analysis of vehicular

communication networks [27]. As a potential theoretical application of these two statistics, we could then investigate a proof of the fact that σ_k never converges to a Poisson distribution, as $\|x - y\| \rightarrow \infty$, via the Chen-Stein method for Poisson approximation [28], [29], since this requires only the first two moments of σ_k to be calculated. But here we restrict our attention to detailing the exact formulas for $\text{Var}(\sigma_k)$ for $k \leq 3$, and $\mathbb{E}\sigma_k$ for all k . We use a form of the Slivnyak-Mecke formula [30] from stochastic geometry [31], [32]. We also detail the specific case of Rayleigh fading as a practically important example. Finally, we note that these results may also have relevance to continuum percolation, since in e.g. the Boolean model [33], the existence of at least one k -hop path from the origin of e.g. the complex plane ‘to infinity’ is of interest [34]–[37].

This paper is structured as follows. In Section II we introduce the random connection model, and summarise our main results. In the following two sections, we then derive the mean and variance for the non-trivial case $k = 3$, and the mean also for general $k \in \mathbb{N}$, showing how the variance is a sum of four terms, all of which are dominated in the limit by a single term equal to the expectation. We also provide numerical corroboration, including of an approximation to the probability that there exist zero k -hop paths for each $k \in \mathbb{N}$. We then discuss the results in relation to path existence in Section V, and conclude in Section VI.

II. SUMMARY OF MAIN RESULTS

The *Random Connection Model* is a graph $\mathcal{G}_H = (\mathcal{V}, E)$ formed on a random subset \mathcal{V} of \mathbb{R}^d by adding an edge between distinct pairs of \mathcal{V} with probability $H(\|x - y\|)$, where $H : \mathbb{R}^+ \rightarrow [0, 1]$ is called the *connection function*, and $\|x - y\|$ is Euclidean distance. Often \mathcal{V} is a Poisson point process of intensity ρdx , with dx Lebesgue measure on \mathbb{R}^d . By *k-hop path* we mean a non-repeating sequence of k adjacent edges joining two different vertices x, y in the vertex set of \mathcal{G}_H . Since we only add edges between distinct pairs of \mathcal{V} , vertices do not connect to themselves in what follows. This forbids paths of two hops becoming three hops simply by connecting vertices to themselves at some point along the path. See e.g. Fig. 1, which shows an example case for $k = 3$. We also consider the practically important case of Rayleigh fading [38] where, with $\beta > 0$ a parameter and $\eta > 0$ the path loss exponent, the connection function, with $\|x - y\| > 0$, is given by

$$H(\|x - y\|) = \exp(-\beta\|x - y\|^\eta) \quad (3)$$

and is otherwise zero. This choice is discussed in e.g. Section 2.3 of [7]. Note that we refer to *nodes* when discussing actual communication devices in a wireless network, and *vertices* when discussing their associated graphs directly. We now detail our main results.

Theorem II.1 (Expected number of k -hop paths for general H). *Take a general connection function $H : \mathbb{R}^+ \rightarrow [0, 1]$. Define a new Poisson point process \mathcal{Y}^* which is \mathcal{Y} conditioned on containing two specific points $x, y \in \mathbb{R}^d$ at Euclidean distance $\|x - y\|$. Consider those two vertices x, y in the vertex set of the random geometric graph $\mathcal{G}_H = (\mathcal{Y}^*, E)$, and set $x = z_0, y = z_k$. Then, in \mathcal{G}_H , the expected number of distinct non-repeating sequences of k sequential edges starting at x and terminating at y is*

$$\mathbb{E}\sigma_k = \rho^{k-1} \int_{\mathbb{R}^{dk-d}} dz_1 \dots dz_{k-1} \prod_{i=0}^{k-1} H(\|z_i - z_{i+1}\|). \quad (4)$$

Proposition II.2 (Expected number of k -hop paths for the specific case of Rayleigh fading). *Take H from Eq. 3, and $d, \eta = 2$ so that we consider free space propagation in two dimensions. Define a new Poisson point process \mathcal{Y}^* as in Theorem II.1, and the random geometric graph $\mathcal{G}_H = (\mathcal{Y}^*, E)$. Then, in \mathcal{G}_H , the expected number of distinct non-repeating sequences of k sequential edges starting at x and terminating at y is*

$$\mathbb{E}\sigma_k = \frac{1}{k} \left(\frac{\rho\pi}{\beta} \right)^{k-1} \exp \left(\frac{-\beta\|x - y\|^2}{k} \right). \quad (5)$$

Theorem II.3 (Variance of the number of three-hop paths for the specific case of Rayleigh fading). *Take H from Eq. 3, and $d, \eta = 2$ so that we consider free space propagation in two dimensions. Define a new Poisson point process \mathcal{Y}^* as in Theorem II.1, and the random geometric graph $\mathcal{G}_H = (\mathcal{Y}^*, E)$. Then, in this graph, the variance of the number of distinct non-repeating sequences of three sequential edges starting at x and terminating at y is*

$$\begin{aligned} \text{Var}(\sigma_3) = \mathbb{E}\sigma_3 \\ + \frac{\pi^3 \rho^3}{\beta^3} \left(\frac{1}{4} \exp \left(\frac{-\beta\|x - y\|^2}{2} \right) + \frac{1}{6} \exp \left(\frac{-3\beta\|x - y\|^2}{4} \right) \right) \\ + \frac{\pi^2 \rho^2}{8\beta^2} \exp(-\beta\|x - y\|^2). \end{aligned} \quad (6)$$

Remark II.4. *For the case $k = 3$, the exponent in Eq. 5 has coefficient $1/3$, which is the least negative of the four terms in Eq. 6. This particular exponential therefore dominates the other terms as $\beta\|x - y\| \rightarrow \infty$, and $\lim_{\beta\|x-y\| \rightarrow \infty} \text{Var}(\sigma_3) = \mathbb{E}\sigma_3$.*

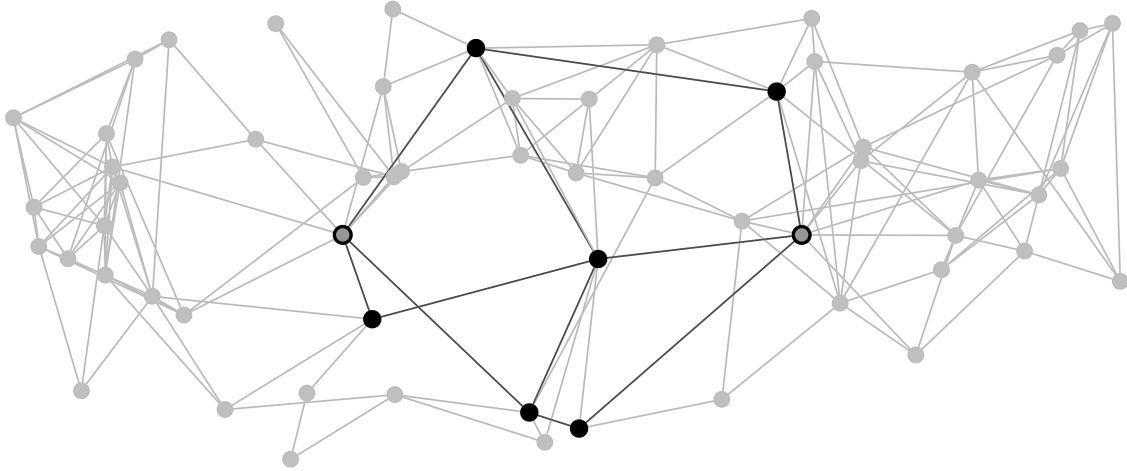


Fig. 1. Example of a soft random geometric graph with connection function taken from Eq. 3. The Euclidean separation between x and y is 3 units taking $\beta = 1$ and an expected $\rho = 1$ node per unit area. Here $\sigma_3 = 5$, $\mathbb{E}\sigma_3 = 2.36$ and $\text{Var}(\sigma_3) = 9.95$, according to Eqs. 5 and 6. All three hop paths between two nodes x and y , distinguished with the thick node borders, are highlighted in black.

III. THE EXPECTED NUMBER OF k -HOP PATHS

Consider a point process \mathcal{X} on some space \mathcal{V} . If it is assumed that $x \in \mathcal{V}$ and $x \in \mathcal{X}$, what is true of the remaining points $\mathcal{X} \setminus \{x\}$? The Poisson point process has the property that when conditioning on a point $x \in \mathcal{V}$ being included in \mathcal{X} , the remaining points $\mathcal{X} \setminus \{x\}$ are still a point process, and of the original intensity. This is *Slivnyak's theorem*, and it characterises the Poisson process, see e.g. Proposition 5 in [39]. The relevance of the following lemma [5], [30] from stochastic geometry [31], [32] is now framed.

Lemma III.1 (Slivnyak-Mecke Formula). *Let $t \in \mathbb{N}$. For any measurable real valued function f defined on the product of $(\mathbb{R}^d)^t \times \mathcal{G}$, where \mathcal{G} is the space of all graphs on finite subsets of $[0, 1]^d$, given a connection function H , the following relation holds*

$$\begin{aligned} \mathbb{E} \sum_{X_1, \dots, X_t \in \mathcal{V}}^{\neq} f(X_1, \dots, X_t, \mathcal{G}_H(\mathcal{Y} \setminus \{X_1, \dots, X_t\})) \\ = n^t \int_{[0,1]^d} dx_1 \cdots \int_{[0,1]^d} dx_t \mathbb{E} f(x_1, \dots, x_t, \mathcal{G}_H(\mathcal{Y})) \end{aligned} \quad (7)$$

where $\mathcal{Y} \subset [0, 1]^d$, $\mathbb{E}\|\mathcal{Y}\| = n$, and \sum^{\neq} means the sum over all ordered t -tuples of distinct points in \mathcal{Y} .

Remark III.2. *To clarify, note that $\{a, b\}$ and $\{b, a\}$ are distinct ordered 2-tuples, but indistinct unordered 2-tuples.*

Proof. In the case $t = 2$ with

$$f(u, v, \mathcal{G}_H(\mathcal{Y})) =: \mathbf{1}\{u \leftrightarrow v\} \quad (8)$$

a Bernoulli variate with parameter $H(\|u - v\|)$, then

$$\mathbb{E}f(u, v, \mathcal{G}_H(\mathcal{Y})) = H(\|u - v\|) \quad (9)$$

where the expectation is over all graphs $\mathcal{G}_H(\mathcal{Y})$. These indicator functions are important for dealing with the existence of edges between points of \mathcal{Y} . We note this for clarity, but it is not required for what follows. The proof of Lemma III.1 is obtained by conditioning on the number of points of \mathcal{Y} . Firstly,

$$\begin{aligned} \mathbb{E} \sum_{X_1, \dots, X_m \in \mathcal{Y}}^{\neq} f(X_1, \dots, X_m, \mathcal{G}_H(\mathcal{Y} \setminus \{X_1, \dots, X_m\})) &= \sum_{t=m}^{\infty} \left(\frac{e^n n^t}{t!} \right) (t)_m \int_{[0,1]^d} dx_1 \dots \\ &\dots \int_{[0,1]^d} dx_t f(x_1 \dots x_m, \mathcal{G}_H(\{x_{m+1}, \dots, x_t\})) \end{aligned} \quad (10)$$

where $(n)_k = n(n-1)\dots(n-k-1)$ is the descending factorial. Bring the m -dimensional integral over positions of vertices in the m -tuple outside the sum,

$$\begin{aligned} n^m \int_{[0,1]^d} dx_1 \dots \int_{[0,1]^d} dx_m \sum_{t=m}^{\infty} \left(\frac{e^n n^{t-m}}{(t-m)!} \right) \\ \times \int_{[0,1]^d} dy_1 \dots \int_{[0,1]^d} dy_{t-m} f(x_1 \dots x_m, \mathcal{G}_H(\{y_1, \dots, y_{t-m}\})), \end{aligned}$$

and change variables such that $r = t - m$, such that

$$\begin{aligned} n^m \int_{[0,1]^d} dx_1 \dots \int_{[0,1]^d} dx_m \sum_{r=0}^{\infty} \left(\frac{e^n n^r}{r!} \right) \int_{[0,1]^d} dy_1 \int_{[0,1]^d} dy_r f(x_1 \dots x_m, \mathcal{G}_H(\{y_1, \dots, y_r\})) \\ = n^m \int_{[0,1]^d} dx_1 \dots \int_{[0,1]^d} dx_m \mathbb{E}f(x_1, \dots, x_m, \mathcal{G}_H(\mathcal{Y})) \end{aligned}$$

as required. \square

We now provide a general formula for the expected number of k -hop paths between $x, y \in V$.

Proof of Theorem II.1. Define a new Poisson point process \mathcal{Y}^* conditioned on containing two specific points $x, y \in \mathbb{R}^d$ at Euclidean distance $\|x - y\|$ and set $x = z_0, y = z_k$. In a similar manner to Eq. 8, define the *path-existence function* g to be the following product

$$g(z_1, \dots, z_{k-1}, \mathcal{G}_H(\mathcal{Y}^*)) = \prod_{i=0}^{k-1} \mathbf{1}\{z_i \leftrightarrow z_{i+1}\} \quad (11)$$

where the indicator is defined in Eq. 8. The expected value of this function is then just the product of the connection probabilities H of the inter-point distance along the sequence z_0, \dots, z_k , i.e.

$$\mathbb{E}g(z_1, \dots, z_{k-1}, \mathcal{G}_H(\mathcal{Y}^*)) = \prod_{i=0}^{k-1} H(\|z_i - z_{i+1}\|) \quad (12)$$

From the Mecke formula

$$\mathbb{E} \sum_{X_1, \dots, X_{k-1} \in \mathcal{Y}^*}^{\neq} g(X_1, \dots, X_{k-1}) = \rho^{k-1} \int_{\mathbb{R}^{2k-2}} \mathbb{E}g(z_1, \dots, z_{k-1}) dz_1 \dots dz_{k-1} \quad (13)$$

and with Eq. 12 replacing the integrand on the right hand side, the proposition follows. \square

We now expand on the practically important situation where vertices connect with probability given by Eq. 3.

Proof of Proposition II.2. Using Eq. 13 with H taken from Eq. 3, and in the case $d, \eta = 2$, we have

$$\begin{aligned} \mathbb{E} \sum_{X_1, \dots, X_{k-1} \in \mathcal{Y}^*}^{\neq} g(X_1, \dots, X_{k-1}, \mathcal{G}_H(\mathcal{Y}^* \setminus X_1, \dots, X_{k-1})) \\ = \rho^{k-1} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} dz_{1_x} dz_{1_y} \dots dz_{(k-1)_x} dz_{(k-1)_y} \\ \times \exp \left(-\beta \left(z_{1_x}^2 + \dots + \left(\|x - y\| - z_{(k-1)_x}^2 \right) + z_{(k-1)_y}^2 \right) \right) \end{aligned}$$

which, by repeatedly integrating, demonstrates Eq. 5. \square

IV. THE VARIANCE FOR $k = 3$

In this section we consider the variance of the number of paths of three sequential edges.

Proof of Theorem II.3. A similar technique to the one implemented here is used to derive the asymptotic variance of the number of edges in the random geometric graph G defined in Section II, see e.g. Section 2 of [30].

The proof now follows. Consider $\mathbb{E}\sigma_3^2(\|x - y\|)$. This is the expected number of ordered *pairs* of three hop paths between the fixed vertices x and y . There are three non-overlapping contributions,

$$\sigma_3^2 = \Sigma_0 + \Sigma_1 + \Sigma_2, \quad (14)$$

where for $i = 0, 1, 2$ the integer Σ_i denotes the number of ordered pairs of three hop paths with i vertices in common. Taking g from Eq. 11, we can quickly evaluate the term Σ_0 , which is the following sum over ordered 4-tuples of points in \mathcal{Y}^* ,

$$\Sigma_0 = \sum_{V, W, X, Y \in \mathcal{Y}^*}^{\neq} g(V, W) g(X, Y). \quad (15)$$

The Mecke formula implies that

$$\mathbb{E}\Sigma_0 = \rho^4 \int_{\mathbb{R}^8} \mathbb{E}(g(z_1, z_2) g(z_3, z_4)) dz_1 dz_2 dz_3 dz_4, \quad (16)$$

and since, according to Eq. 13, we have

$$\mathbb{E}\sigma_3 = \rho^2 \int_{\mathbb{R}^4} \mathbb{E}g(z_1, z_2) dz_1 dz_2 \quad (17)$$

then $\mathbb{E}\Sigma_0 = (\mathbb{E}\sigma_3)^2$, which cancels with a term in the definition of the variance $\text{Var}(\sigma_3) = \mathbb{E}(\sigma_3^2) - (\mathbb{E}(\sigma_3))^2$, such that we have the following simpler expression for the variance, based on Eq. 14 and Eq. 16,

$$\text{Var}(\sigma_3) = \mathbb{E}\Sigma_1 + \mathbb{E}\Sigma_2. \quad (18)$$

Now, $\text{Var}(\sigma_3)$ will follow from a careful evaluation of $\mathbb{E}\Sigma_1$ and $\mathbb{E}\Sigma_2$. The first of these can be broken down into two separate contributions, denoted $\mathbb{E}\Sigma_{1(1)}$ and $\mathbb{E}\Sigma_{1(2)}$. For $\mathbb{E}\Sigma_{1(1)}$, notice the left panel of Fig. 2, which shows an intersecting pair of paths in $\mathcal{G}_H(\mathcal{Y}^*)$ which share a single vertex U which is itself connected by an edge to y . Many triples of points in $\mathcal{Y}^* \setminus \{x, y\}$ display this property. Notice that

$$\Sigma_{1(1)} = \sum_{U \in \mathcal{Y}^*} \sum_{W, Z \in \mathcal{Y}^* \setminus \{U\}}^{\neq} g(U, W) g(U, Z), \quad (19)$$

which, via the Mecke formula, and with U the position vector of the shared vertex, gives

$$2\rho \int_{\mathbb{R}^2} H(\|y - U\|) \mathbb{E} \left[\left(\sum_{X \in \mathcal{Y}^* \setminus \{U\}} g(X) \right)_2 \right] dU \quad (20)$$

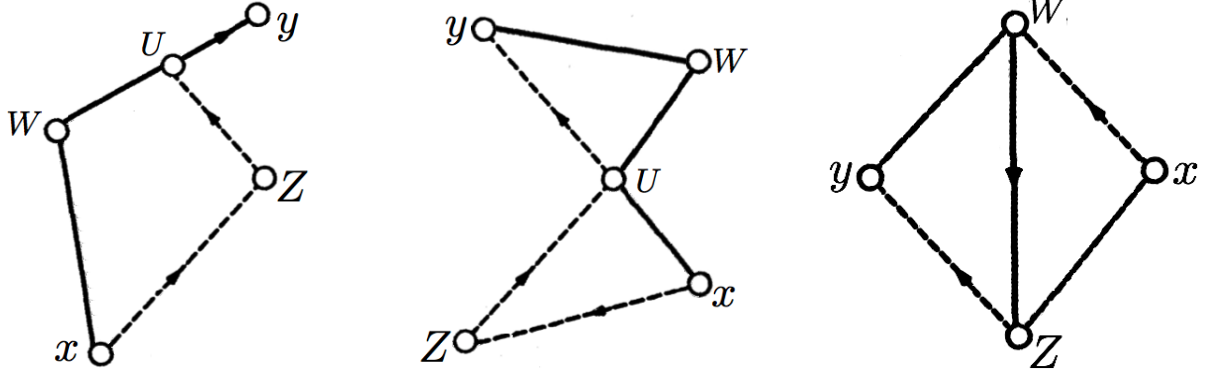


Fig. 2. *Left:* A pair of paths from x to y in the random connection model which intersect at exactly one vertex U , in a specific way indicated by the diagram, i.e. they share only the edge from U to y . *Center:* The same event, but occurring in such a way that they share no edges. These motif-like objects are used to calculate the variance of the number of three-hop paths, since one can represent the statistic as the number of ordered *pairs* of paths running between two points. *Right:* Two three-hop paths from x to y intersecting at both their vertices. They share the single edge from W to Z .

with $(a)_2 = a(a-1)$. Note the factor of two is required because there exist a *pair* of non-isomorphic structures which can form from paths intersecting in the manner of the left panel of Fig. 2. To see the other, just swap the labels of x and y . To expand on Eq. 20 slightly, for either choice of non-isomorphic structures just discussed, count all two-hop paths from x to U , as in the diagram, and count all ordered pairs of these paths. Each corresponds to a distinct, ordered pair of three-hop paths *so long as U then actually connects to y* . As an example, consider the situation where there are three two-hop paths from x to U . There would then be $(3)_2 = 6$ distinct ordered pairs of three-hop paths from x and y which each only intersect at U should U be joined by an edge to y . Also, be careful not to miss the necessary factor of two in Eq. 20.

Since the descending factorial in Eq. 20 is just $(\sigma_2(\|x - U\|))_2$, and briefly noting that with $\Pi \sim \text{Po}(\lambda)$ then $\mathbb{E}(\Pi)_2 = \lambda^2$, we have part of the integrand of Eq. 20,

$$\mathbb{E} \left[\left(\sum_{X \in \mathcal{Y}^* \setminus \{U\}} f(X) \right)_2 \right] = (\mathbb{E} \sigma_2(\|x - U\|))^2. \quad (21)$$

By multiplying Eq. 21 by $H(\|y - U\|)$ and integrating over all feasible positions of U , $\Sigma_{1(1)}$ will follow. Eq. 20 thus becomes

$$2\rho \int_{\mathbb{R}^2} H(\|y - U\|) (\mathbb{E} \sigma_2(\|x - U\|))^2 dU \quad (22)$$

and so $\Sigma_{1(1)}$ is

$$\Sigma_{1(1)} = 2\rho^3 \int_{\mathbb{R}^2} H(\|y - U\|) \times \left(\int_{\mathbb{R}^2} H(\|x - z\|) H(\|z - U\|) dz \right)^2 dU \quad (23)$$

in terms of a general connection function, and for the case of Rayleigh fading taking $d, \eta = 2$, Eq. 23 evaluates to

$$\Sigma_{1(1)} = \frac{\pi^3 \rho^3}{4\beta^3} \exp\left(\frac{-\beta\|x - y\|^2}{2}\right), \quad (24)$$

which appears as the second term in Eq. 6.

Now consider $\Sigma_{1(2)}$. This counts pairs of paths which share a single vertex, but in a different way. This is depicted in the middle panel of Fig. 2. Paths are paired with each other when one takes a direct route to U , while the other an indirect route via Z . As before, they can be counted via the Mecke formula,

$$\Sigma_{1(2)} = \sum_{U \in \mathcal{Y}^*} \sum_{Z \in \mathcal{Y}^* \setminus \{W\}} g(Z, U) \sum_{W \in \mathcal{Y}^* \setminus \{Z\}} g(U, W). \quad (25)$$

In a similar manner to the evaluation of $\Sigma_{1(1)}$, the two inner sums are in fact just counting the number of two hop paths between x and U , and U and y , then pairing them with each other by taking the product of one count with the other. From the Mecke formula, $\mathbb{E}\Sigma_{1(2)}$ can be written

$$\rho \int_{\mathbb{R}^2} H(\|x - U\|) H(\|U - y\|) \mathbb{E} \left[\sum_{Z \in \mathcal{Y}^* \setminus \{W\}} g(Z, U) \right] \mathbb{E} \left[\sum_{W \in \mathcal{Y}^* \setminus \{Z\}} g(U, W) \right] dU \quad (26)$$

since the two sums are independent. By writing these expectations as $\mathbb{E}(\|x - U\|)$ and $\mathbb{E}(\|U - y\|)$, noting this time there is no squaring or descending factorials as we simply take the product of the two variables, the right hand side of Eq. 26 simplifies to

$$\rho \int_{\mathbb{R}^2} H(\|x - U\|) H(\|U - y\|) \mathbb{E}(\|x - U\|) \mathbb{E}(\|U - y\|) dU, \quad (27)$$

and $\Sigma_{1(2)}$ in terms of a general connection function is therefore

$$\begin{aligned} \mathbb{E}\Sigma_{1(2)} = \rho^3 \int_{\mathbb{R}^2} H(\|x - U\|) H(\|U - y\|) & \left(\int_{\mathbb{R}^2} H(\|x - z\|) H(\|z - U\|) dz \right) \\ & \times \left(\int_{\mathbb{R}^2} H(\|U - z\|) H(\|z - y\|) dz \right) dU \end{aligned} \quad (28)$$

and for the case of Rayleigh fading taking $d, \eta = 2$, the third term on the right hand side of Eq. 6 is

$$\mathbb{E}\Sigma_{1(2)} = \frac{1}{6} \exp\left(\frac{-3\beta\|x - y\|^2}{4}\right) \quad (29)$$

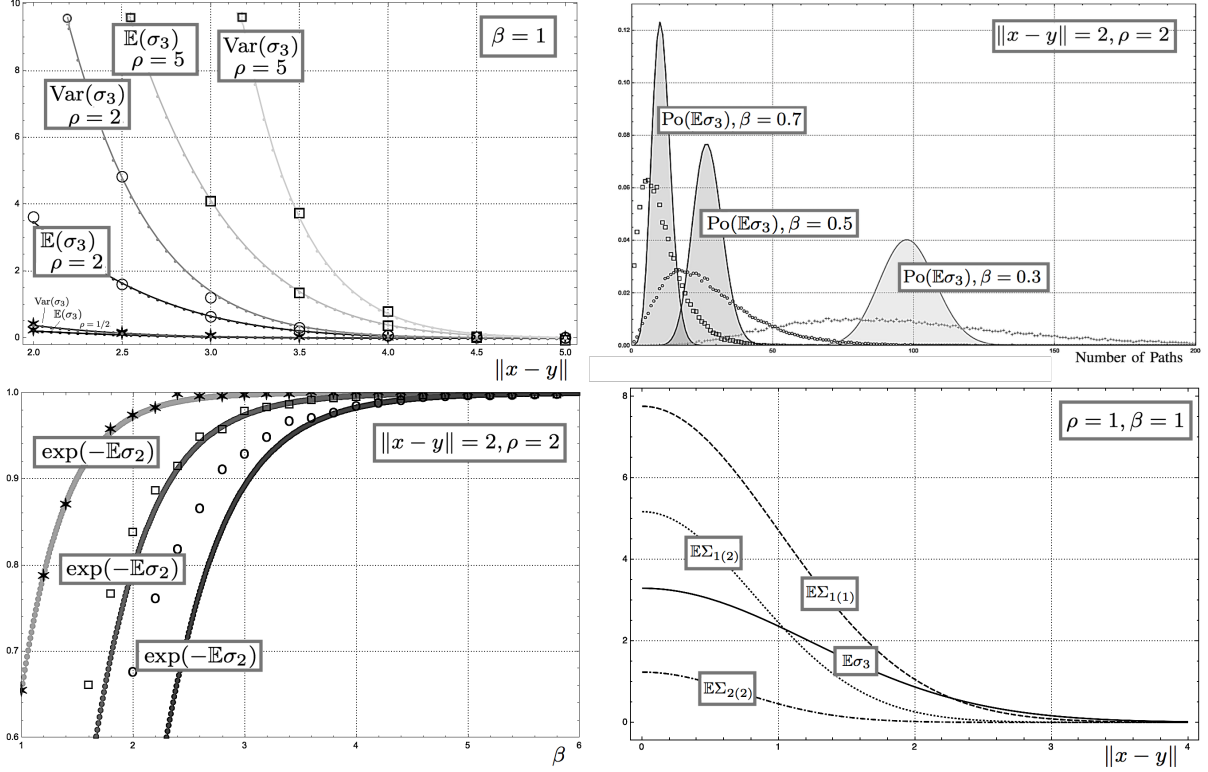


Fig. 3. *Top left*: Numerical corroboration of the analytic mean and variance of σ_3 , the number of three-hop paths joining two vertices x, y in \mathcal{G}_H , for three separate densities $\rho = 1/2, 2$ and 5 , taking $\beta = 1$, scaling over Euclidean separation (horizontal axis). The \star , \circ and \square 's are Monte Carlo data, averaged over 10^4 random graphs, whereas the smooth lines are our equations found in Theorems II.2 and II.3. *Top right*: The distribution of the number of k -hop paths in the random connection model, fixing $\|x-y\|, \rho = 2$, for three values of $\beta = 0.7, 0.5, 0.3$ (left to right). The variance exceeds the mean to a greater extent as the typical connection range grows at fixed density. The Poisson distribution with the analytic mean of Eq. 5 is plotted for each value of β for comparison. *Bottom left*: The probability that there exists zero k -hop paths $\exp(-\mathbb{E}\sigma_k)$ for $k = 2, 3, 4$ (left to right), compared with numerical data (shapes), ranging over increasing β (horizontal axis). The approximation deviates from the numerical data where there is a noticeable difference between the mean and variance. *Bottom right*: The four terms in Eq. 6 are compared at fixed $\rho, \beta = 1$ as $\|x-y\|$ increases. Note $\mathbb{E}\Sigma_{2(1)} = \mathbb{E}\sigma_3$, according to Eq. 30. Other than the mean, the $\mathbb{E}\Sigma_{1(1)}$ remains the only non-negligible contribution in the large-distance limit.

by integrating the product of exponentials.

There are two more terms, $\mathbb{E}\Sigma_{2(1)}$ and $\mathbb{E}\Sigma_{2(2)}$. These both correspond to pairs of paths which share two vertices. Firstly, $\Sigma_{2(1)}$ refers to pairs of paths which share two vertices and all their edges, and so

$$\mathbb{E}\Sigma_{2(1)} = \mathbb{E}\sigma_3 \quad (30)$$

since there is a pair of paths for each path, specifically the self-pair. Secondly, $\Sigma_{2(2)}$ refers to

pairs which share all their vertices, but not all their edges. This pairing is depicted in the right panel of Fig 2. For this term, the Mecke formula gives

$$\Sigma_{2(2)} = \sum_{Z, W \in \mathcal{Y}^*} g(Z, W) g(W, Z) \quad (31)$$

and so, in a similar manner to the other terms,

$$\begin{aligned} \mathbb{E}\Sigma_{2(2)} = \rho^2 \int_{\mathbb{R}^2} H(\|x - Z\|) H(\|Z - W\|) H(\|W - y\|) H(\|x - W\|) \\ \times H(\|Z - y\|) dW dZ. \end{aligned} \quad (32)$$

Only once all five links form do these pairs appear, so they are rare, and $\mathbb{E}\Sigma_{2(2)}$ is relatively small. Note that extensive counts of pairs of paths with this property can be an indication of proximity, a point we expand upon in Section V. Finally, therefore, the last term in Eq. 6 is

$$\Sigma_{2(2)} = \frac{\pi^2 \rho^2}{8\beta^2} \exp(-\beta\|x - y\|^2) \quad (33)$$

via evaluating Eq. 32, and via Eq. 18 and then Eqs. 24, 29, 30 and 33, the theorem follows. \square

We numerically corroborate Theorems II.2 and II.3 via Monte Carlo simulations, the results of which are presented in the top-left panel of Fig. 3.

V. DISCUSSION

A. Composite variance

Consider the lower-right panel of Fig. 3. All four terms in Eq. 6 are compared as $\|x - y\|$ grows, fixing ρ, β . All the terms decay to zero, and the expectation eventually dominates. Only $\Sigma_{2(1)}$ is non-negligible in the large-distance limit. We are not yet able to conceive of a way to use this information to approximate the probability a k -hop path exists, beyond simply $P(\sigma_k = 0) \leq \exp(-\mathbb{E}\sigma_k)$, see the lower-left panel of Fig. 3, but it would appear that there is a ‘right’ way to do this which does not ignore the spatial dependence, given some limit is taken. Note also that the four-termed variance of Eq. 6 is not specific to a Rayleigh fading model, but will be a property of the statistic for any choice of connection function.

B. Scaling of statistics with density and connection range

With $\beta, \|x - y\|$ fixed, the expected number of paths is $\mathcal{O}(\rho^k)$, while the variance appears to be $\mathcal{O}(\rho^{k+1})$. We have only verified this for $k = 3$. We numerically obtain the probability mass of

σ_3 in the top-right panel of Fig. 3 by generating 10^5 random graphs and counting all three hop paths between two extra vertices added at fixed distance $\|x - y\|$ taking $\rho = 2$ for $\beta = 0.7, 0.5$ and 0.3 . Plotted for comparison is the mass of a Poisson distribution with mean given by Eq. 5. This highlights the excessive dispersion of the data, which is particularly prominent for the case $\beta = 0.3$. In fact, the distribution is so disperse that it appears practically flat around its mean. This may allow a uniform approximation.

C. Normal approximation and localisation

A normal approximation $\sigma \sim \mathcal{N}(\mathbb{E}\sigma_k, \text{Var}(\sigma_k))$ may be appropriate when the network is extremely dense, such as in a swarm robotics scenario. With the mean and variance provided, the exact distribution in the *dense limit* where $\rho(n) \int_{\mathbb{R}^d} H(z) dz \rightarrow \infty$ as $n \rightarrow \infty$ (the expected degree of a typical vertex diverges) can be described precisely by

$$\frac{\sigma_k - \mathbb{E}\sigma_k}{\sqrt{\text{Var}(\sigma_k)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) \quad (34)$$

though this central limit theorem has not been proven, see e.g. Section 3 of [30] for a discussion of how this is done when considering instead the random number of edges (in the whole graph) in the deterministic case. Only at extreme densities beyond those anticipated of ultra-dense deployment is such an approximation accurate enough to use [16]. It may, however, be possible to approximate the distribution of σ_k with the negative binomial mass function [40] with the derived mean and variance, which could prove useful in applications where the exact distribution is required, since the negative binomial mass function is characteristic of other scenarios where statistical dependence in space or time is apparent. An example is the clustering in time of tropical cyclones, since one storm essentially encourages another according to a sophisticated meteorological process [41], [42].

Localisation is an example where details of each probability would prove useful, since one could use them in an attempt to range the distance to a target node by referencing the number of e.g. three hop paths, ascertained from connectivity information, with a conditional distribution $P(i \text{ paths of } k \text{ hops} \mid \text{Euclidean distance } \delta)$, providing a more sensitive distance estimate than via DV-hop [43], or extensions which utilise the distribution of the number of hops itself [21], since small movements can have a dramatic effect on the path count. In extremely dense networks, however, this may not work since the excessive variance of the number of paths provides many candidate destinations of equal likelihood given a path count ascertained from

connectivity information. Despite this uncertainty, one may be able to obtain noisy *internode* distance estimates using only the network's connectivity information, replacing the need for ultra-wideband (UWB) sensors in certain scenarios. These internode distances can be entered into a combinatorial optimisation problem whose solution is a set of device locations [44], [45]. The advantage is of course that only connectivity information in the form of a simple adjacency matrix is required, rather than actively ranging with time-difference-of-arrival (TDOA), or angle-of-arrival etc, and moreover, that spatial dependence is considered a sort of noise, though the effect of this has not been studied.

Finally, we note that the path count could be replaced with a more sophisticated topological measure, perhaps based on a form of eigenvector centrality (see e.g. [46]–[48] for a discussion of this topic in the scope of multi-hop communication), since two nearby nodes have highly correlated eigenvector centrality indices since they share neighbours (spatial dependence), and so this correlation can be an indication and measure of closeness in space.

VI. CONCLUSION

In a random geometric graph known as the random connection model, we derived both the mean and variance of the number of k -hop paths between two nodes x, y at displacement $\|x - y\|$, on condition that $k \in [1, 3]$. We also provided details of an example case whenever Rayleigh fading statistics are observed, which is important in applications. This shows how the variance of the number of paths is in fact composed of four terms, no matter what connection function is used. This provides an approximation to the probability that a k -hop path exists between distant vertices, and hints at a technique for formulating the ‘right’ approximation in general, which is a sort of correction to a mean field model. This works toward addressing a recent problem of Mao and Anderson. It also can be used to experiment with connectivity based localisation, such as estimating internode distances without ranging with ultra-wideband sensors.

ACKNOWLEDGEMENTS

This work is supported by Samsung Research Funding and Incubation Center of Samsung Electronics under Project Number SRFC-IT-1601-09. The first author also acknowledges support from the EPSRC Institutional Sponsorship Grant 2015 *Random Walks on Random Geometric Networks*, and wishes to thank Mathew Penrose, Kostas Koufos, David Simmons, Georgie Knight, Orestis Georgiou, Carl Dettmann, Justin Coon and Leo Laughlin for helpful discussions.

REFERENCES

- [1] M. D. Penrose, "The Longest Edge of the Random Minimal Spanning Tree," *The Annals of Applied Probability*, vol. 7, no. 2, pp. 340–361, 1997.
- [2] P. Gupta and P. R. Kumar, "Critical Power for Asymptotic Connectivity," in *Proceedings of the 37th IEEE Conference on Decision and Control*, 1998, pp. 1106–1110.
- [3] M. Walters, "Random Geometric Graphs," in *Surveys in Combinatorics 2011*, Robin Chapman, Ed. Cambridge University Press, 2011.
- [4] M. Penrose, "Random graphs: The Erdős-Rényi model," in *Talk at Oberwolfach Workshop on Stochastic Analysis for Poisson Point Processes, February 2013, available at the homepage of the author*, 2013.
- [5] M. D. Penrose, "Connectivity of Soft Random Geometric Graphs," *The Annals of Applied Probability*, vol. 26, no. 2, pp. 986–1028, 2016.
- [6] J. Coon, C. Dettmann, and O. Georgiou, "Full Connectivity: Corners, Edges and Faces," *The Journal of Statistical Physics*, vol. 147, no. 4, pp. 758–778, 2012.
- [7] A. P. Giles, O. Georgiou, and C. P. Dettmann, "Connectivity of Soft Random Geometric Graphs over Annuli," *The Journal of Statistical Physics*, vol. 162, no. 4, p. 10681083, 2016.
- [8] M. D. Penrose, *Random Geometric Graphs*. Oxford University Press, 2003.
- [9] S. K. Iyer, "Connecting the Random Connection Model," *preprint arXiv:1510.05440*, 2015.
- [10] G. Mao, Z. Zhang, and B. Anderson, "Probability of k-Hop Connection under Random Connection Model," *Communications Letters, IEEE*, vol. 14, no. 11, pp. 1023–1025, 2010.
- [11] G. Mao and B. D. Anderson, "On the Asymptotic Connectivity of Random Networks under the Random Connection Model," *INFOCOM, Shanghai, China*, p. 631, 2011.
- [12] G. Mao and B. D. O. Anderson, "Connectivity of large wireless networks under a general connection model," *IEEE Transactions on Information Theory*, vol. 59, no. 3, pp. 1761–1772, 2013.
- [13] P. Gupta and P. R. Kumar, "The Capacity of Wireless Networks," *IEEE Transactions on Information Theory*, vol. 42, no. 6, pp. 388–404, 2000.
- [14] S. Dulman, M. Rossi, P. Havinga, and M. Zorzi, "On the hop count statistics for randomly deployed wireless sensor networks," *Int. J. Sen. Netw.*, vol. 1, no. 1/2, pp. 89–102, Sep. 2006.
- [15] X. Ta, G. Mao, and B. Anderson, "On the Probability of K-hop Connection in Wireless Sensor Networks," *IEEE Communications Letters*, vol. 11, no. 9, 2007.
- [16] X. Ge, S. Tu, G. Mao, C. X. Wang, and T. Han, "5G Ultra-Dense Cellular Networks," *IEEE Wireless Communications*, vol. 23, no. 1, pp. 72–79, 2016.
- [17] S. C. Ng and G. Mao, "Analysis of k-hop connectivity probability in 2-d wireless networks with infrastructure support," in *2010 IEEE Global Telecommunications Conference*, 2010, pp. 1–5.
- [18] A. Zemlianov and G. de Veciana, "Capacity of ad hoc wireless networks with infrastructure support," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 3, pp. 657–667, 2005.
- [19] C. Bettstetter and J. Eberspacher, "Hop distances in homogeneous ad hoc networks," in *The 57th IEEE Semiannual Vehicular Technology Conference*, vol. 4, 2003, pp. 2286–2290.
- [20] E. Perevalov, R. Blum, A. Nigara, and X. Chen, "Route discovery and capacity of ad hoc networks," in *IEEE Global Telecommunications Conference, 2005.*, 2005.
- [21] C. Nguyen, O. Georgiou, and Y. Doi, "Maximum Likelihood Based Multihop Localization in Wireless Sensor Networks," in *Proceedings. IEEE ICC, London, UK*, 2015.

- [22] J. Díaz, D. Mitsche, G. Perarnau, and X. Pérez-Giménez, "On the relation between graph distance and euclidean distance in random geometric graphs," *Advances in Applied Probability*, vol. 48, no. 3, p. 848864, 2016.
- [23] P. Clifford, "Markov Random Fields in Statistics," in *Disorder in Physical Systems: A Volume in Honour of John M. Hammersley*, Geoffrey Grimmett and Dominic Welsh, Ed. Cambridge University Press, 1990.
- [24] I. P. Goulden and D. M. Jackson, *Combinatorial Enumeration*. Dover Publications, 2004.
- [25] A. P. Kartun-Giles, G. Knight, O. Georgiou, and C. P. Dettmann, "The electrical resistance of a one-dimensional random geometric graph," *In preparation*, 2017.
- [26] G. Parisi, "Mean field theory for spin glasses," in *Recent Topics in Theoretical Physics: Proceedings of the First Nishinomiya-Yukawa Memorial Symposium, Nishinomiya, Japan, November 8–9, 1986*, H. Takayama, Ed., 1988.
- [27] G. Knight, A. P. Kartun-Giles, O. Georgiou, and C. P. Dettmann, "Counting geodesic paths in 1-d vanets," *IEEE Wireless Communications Letters*, vol. 6, no. 1, pp. 110–113, 2017.
- [28] R. Arratia, L. Goldstein, and L. Gordon, "Two moments suffice for poisson approximations: The chen-stein method," *Ann. Probab.*, vol. 17, no. 1, pp. 9–25, 1989.
- [29] L. H. Y. Chen, "Poisson approximation for dependent trials," *Ann. Probab.*, vol. 3, no. 3, pp. 534–545, 06 1975.
- [30] M. D. Penrose, "Lectures on random geometric graphs," in *Random Graphs, Geometry and Asymptotic Structure*, N. Fountoulakis and D. Hefetz, Eds. Cambridge University Press, 2016, pp. 67–101.
- [31] H. ElSawy, A. Sultan-Salem, M. S. Alouini, and M. Z. Win, "Modeling and analysis of cellular networks using stochastic geometry: A tutorial," *IEEE Communications Surveys Tutorials*, vol. 19, no. 1, pp. 167–203, 2017.
- [32] M. Haenggi, J. Andrews, F. Baccelli, O. Dousse, and M. Franceschetti, "Stochastic geometry and random graphs for the analysis and design of wireless networks," *Selected Areas in Communications, IEEE Journal on*, vol. 7, pp. 1029–1046, 2009.
- [33] M. Franceschetti and R. Meester, *Random Networks for Communication*. Cambridge University Press, 2007.
- [34] S. Broadbent and J. Hammersley, "Percolation Processes," *Proceedings of the Cambridge Philosophical Society*, vol. 63, pp. 629–641, 1957.
- [35] M. D. Penrose, "On a continuum percolation model," *Advances in Applied Probability*, vol. 23, no. 3, pp. 536–556, 1991.
- [36] —, "Continuum percolation and euclidean minimal spanning trees in high dimensions," *The Annals of Applied Probability*, vol. 6, no. 2, pp. 528–544, 05 1996.
- [37] M. Franceschetti, O. Dousse, D. N. C. Tse, and P. Thiran, "Closing the gap in the capacity of wireless networks via percolation theory," *IEEE Transactions on Information Theory*, vol. 53, no. 3, pp. 1009–1018, 2007.
- [38] B. Sklar, "Rayleigh fading channels in mobile digital communication systems. Part I: Characterization," *IEEE Communications Magazine*, vol. 35, no. 7, 1997.
- [39] P. Jagers, "On palm probabilities," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 26, no. 1, pp. 17–32, 1973.
- [40] T. Broderick, L. Mackey, J. Paisley, and M. I. Jordan, "Combinatorial clustering and the beta negative binomial process," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 2, pp. 290–306, 2015.
- [41] H. C. S. Thom, "The distribution of annual tropical cyclone frequency," *Journal of Geophysical Research*, vol. 65, no. 1, pp. 213–222, 1960.
- [42] C. M. Sakamoto, "Application of the poisson and negative binomial models to thunderstorm and hail days probabilities in nevada," *Monthly Weather Review*, vol. 101, no. 4, pp. 350–355, 1973.
- [43] D. Niculescu and B. Nath, "Ad Hoc Positioning System (APS) Using AOA," in *Proceedings of IEEE INFOCOM, San Francisco, CA, USA*, 2003, pp. 1734–1743.

- [44] A. T. Ihler, J. W. Fisher, R. L. Moses, and A. S. Willsky, "Nonparametric belief propagation for self-localization of sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 4, pp. 809–819, 2005.
- [45] N. Kantas, S. S. Singh, and A. Doucet, "Distributed maximum likelihood for simultaneous self-localization and tracking in sensor networks," *IEEE Transactions on Signal Processing*, vol. 60, no. 10, pp. 5038–5047, 2012.
- [46] A. P. Giles, O. Georgiou, and C. P. Dettmann, "Betweenness Centrality in Dense Random Geometric Networks," *Proceedings of IEEE ICC, London, UK*, 2015.
- [47] N. Magaia, A. P. Francisco, P. Pereira, and M. Correia, "Betweenness Centrality in Delay Tolerant Networks," *Ad Hoc Networks*, vol. 33, no. C, pp. 284–305, Oct. 2015.
- [48] A. Jain and B. R. Reddy, "Eigenvector centrality based cluster size control in randomly deployed wireless sensor networks," *Expert Systems with Applications*, vol. 42, no. 5, pp. 2657 – 2669, 2015.